Supplemental material:

Methods: Details of exposure assessment for the SALIA cohort.

Table 1. Comparison of study population and non-responders

Table 2. Association of incident type 2 diabetes with covariates in Cox regression analyses assessing exposure to particulate matter and traffic-related air pollution (as shown in Table 3, main manuscript).

Methods: Details of exposure assessment for the SALIA cohort

Data from monitoring stations in an 8 km grid

Annual means of nitrogen dioxide (NO_2) and airborne particulate matter (total suspended particles; TSP) were obtained from continuous measurements of a telemetric real-time multi-component monitoring system. This monitoring system has been maintained by the State Environment Agency of North Rhine Westphalia since 1979 (www.lanuv.nrw.de/veroeffentlichungen/fachberichte/fachb07/fachbericht7_luanrw.pdf). The measuring sites cover the investigation area in an 8 km grid.

TSP were measured by absorption of beta-radiation of the particulate matter (PM) deposited on a moving band filter (Lahmann 1997). The random cut-off of PM is 15 µm. TSP were converted to PM_{10} by a factor of 0.71. This factor was the mean (range from 0.62 to 0.84) of parallel measurements from 1998 to 2004 at seven measurement stations in the study area of the SALIA cohort (Gehring et al. 2006). NO_2 concentrations were measured by means of chemiluminescence. Air pollution levels measured in Borken (which was close and most comparable to Dülmen) were used for Dülmen, where no monitoring station was available. Air pollution measurements in Borken started in 1990/91 for NO_2 and TSP. Therefore, air pollution concentrations prior to these dates were imputed using air pollution concentrations measured from 1981 to 2000 at 15 monitoring stations in the Ruhr Area assuming similar trends. Likewise the measurements in Essen began in 1989. Values prior to that were imputed by converting values from the neighboring station Essen-Northwest and assuming similar trends (conversion factor 1.14 for TSP and 1.46 for NO_2). Single values missing in the time series of a monitoring site were filled by the mean of the previous and following year. We used mean values of 1986 to 1990 (before the onset of diabetes) to characterize exposure.

We were able to geo-code 97.1% of the participants' baseline addresses. We calculated proximity of homes to major roads, defined as roads with at least 10,000 cars/day, through a geographic information system using traffic count data for the year 2000 provided by the North Rhine-Westphalia State Environment Agency.

Emission inventories in a 1 km grid

Emissions of NO_2 and PM form motor vehicles were available from the State Environment Agency of North Rhine Westphalia (LANUV) for grids of one square kilometre for the year 1994 onward. We used the 1994 values to characterize exposure.

The values are derived from line sources representing the traffic volume of roads of the official categorisation as motorway, federal roads, county roads and rural roads and area sources which represent the remaining road traffic volume specific for different zones as residential, mixed, city and commercial zone. Types of vehicle are categorized in strata: 21 for cars, 5 for light trucks (≤3.5 t) and 5 for heavy trucks (>3.5 t). Factors of emission for each of these strata take into account combustion technology, capacity and age of the vehicle and are defined by the German Federal Environmental Agency (Umweltbundesamt; UBA). In this process the local composition of vehicles as registered in the municipality and its neighbouring zone is taken into account. The exact method for the calculation of these emissions is regulated in the guidelines of VDI (Verein Deutscher Ingenieure, Association of German Engineers) (VDI 2002), emission factors are published by the UBA (see page 9 of http://www.gis.nrw.de/ims/ekatsmall2004/htm/Anleitung_ekl2004_internet.pdf) (Umweltbundesamt 2004).

We assigned the PM and NO_2 emission from the square km in which the participant's address is located, to the individual cohort members. This was possible for all those with geo-codes available.

Individually assigned soot and NO_2 values gained by land-use regression models

Basic measurements were conducted in the one-year period from March 19, 2002 until March 11, 2003. Air samplers were located temporarily at 40 sites distributed north of the Ruhr area including the county of Borken and across Western parts of the Ruhr area. The locations were primarily selected to cover the area of three birth cohort studies (Hochadel et al. 2006). The measurement procedures were performed according to standard operating procedures. More detailed descriptions of these methods have been published elsewhere (Brauer et al. 2003; Cyrys et al. 2003; Hoek et al. 2002).

$PM_{2.5}$ was sampled on Teflon filters with Harvard impactors (Marple et al. 1987). $PM_{2.5}$ absorbance was determined as a marker for black smoke according to ISO 9835. Palmes diffusion tubes (Palmes et al. 1976) were used as passive samplers for NO_2. The absorbed material was analyzed by ion chromatography.

The Gauß-Krüger coordinate system (name of the German geocoding system) was used as spatial reference for the geographical features. A geographical database contained the main road network in the study region. For 14,516 street segments with a total length of 10,329 km, average daily traffic counts from the year 2000 were given. Roads were classified into the official street categorisation: motorway, federal roads, county roads and rural roads. Traffic counts were classified as cars, light trucks (≤3.5 t), heavy trucks (>3.5 t), buses and motor cycles.

Characteristics of the measurement sites and addresses were derived from these raw geographic data by calculating total counts or densities for certain neighbourhoods. For traffic data, circular buffers with radii of 50 m, 100 m, 250 m, 500 m and 1 km, were created around the coordinates of interest and intersected with the road network. The daily traffic flow (DT) within a given buffer was calculated by multiplying the number of vehicles per day with the street length and summing up for all segments in the buffer. The highest number of vehicles per day on a segment in a buffer is the maximum traffic intensity (TI). These values were calculated separately for total vehicles (TV), heavy vehicles (HV) and passenger cars. The distance to the nearest major road and to the nearest highway (motorway or federal road) was determined. In addition, land use data from CORINE (Coordinated Information on the European Environment), obtained from German Remote Sensing Data Center, of the year 2000 was used to calculate areas of different vegetation and agriculture, water, industrial real estate and high density areas in the buffers around the measurement sites.

Annual average concentrations at the measurement sites were derived from the measurement results according to established methods (Brauer et al. 2003). The dependency of the pollutant concentration levels at the forty measurement sites on the traffic-related and population-related characteristics described before was analyzed by multiple linear regression models. The influence of single variables on the fit of the models was explored by stepwise selection methods. We assessed the validity of the models by a cross-validation procedure as described below and by comparing predicted values to independent measurements obtained from the permanent air quality monitoring programme LUQS (Luftqualität-Überwachungssystem) of the Environmental Agency of the state North-Rhine Westphalia (LANUV).

We performed a cross-validation procedure and calculated estimates of the error variance by leaving out each site in turn, the model was fitted with the remaining sites, and the difference of the estimate generated by this model for the left out site and the measured value at this site was calculated. The average of the squared differences is called the cross-validation error (CVE).

Stepwise regression included the predictor variable with highest explained variation (r^2) in every step. This variable was determined a priori by univariate regression analyses with all possible predictors. All remaining variables entered the model without restriction to number of regressors. Regressors with direction of estimate opposite to the expected effect or with less than 1% additional r^2 and regressors which changed the direction of the estimate of regressors included in previous steps were removed during stepwise procedure.

These methods resulted in following equations:

Equation for NO_2 [µg/m^2]: 20.3 + 12.03 address in Ruhr basin + 13.70 daily traffic flow [1000 km/d] of light trucks in 250 m buffer + 23.10 daily traffic flow [1000 km/d] of buses in 250 m buffer – 18.99×10^{-6} forest [m^2] in 500 m buffer + 17.86×10^{-6} high density area [ca] in 250 m buffer.

Equation for $PM_{2.5}$ absorbance: $PM_{2.5}$ [10^{-5}/m]=1.46 + 0.30 address in Ruhr Basin - 54.16×10^{-6} distance [m] to next highway + 0.012 daily traffic flow [1000 km/d] of passenger cars in 250 m buffer + 0.00058 maximal traffic intensity [d^{-1}] of light trucks in 50 m buffer + 34.01×10^{-6} maximal traffic intensity [d^{-1}] of heavy trucks in 500 m buffer + 1.02×10^{-6} high density area [m^2] in 250 m buffer + 4.8×10^{-7} industrial real estate [m^2] in 500 m buffer.

The models could explain 92.1% (adjusted 90.9%) and 86.5% (adjusted 83.6%) of the variance of the measurements for NO_2 and for $PM_{2.5}$ absorbance. RMSE of regression of NO_2 was 2.24 without and 2.67 with cross-validation, RMSE of $PM_{2.5}$ was 0.13 without and 0.19 with cross-validation.

Both regression models showed normal distributed residuals ($p>0.15$), and no heteroscedasticity (White Test: $p=0.128$ for NO_2, $p=0.345$ for $PM_{2.5}$).

Because the one-year measurement programme was not conducted in the whole Ruhr area we investigated the concordance of modelled NO_2 values with measurements of the government covering the whole area in an 8-km grid. We modelled NO_2 values for all addresses of measurement stations and compared these values to the actually measured values. The correlation with annual mean values of measurements done in 2002/2003 was 0.77 for all stations in North-Rhine Westphalia (NRW) (n=47), and 0.59 for stations near the women's places of residence (n=16). Correlations with measurements conducted in 1991 were 0.78 for all stations in NRW (n=74) and 0.61 for stations near the women's place of residence (n=31). This demonstrates good concordance and also a high stability of the pollution pattern over time. Modelled NO_2 and $PM_{2.5}$ absorbance ("soot") were applied to the addresses of the cohort members in the year 1990 to describe individual levels of exposure.

References

Brauer M, Hoek G, van Vliet P, Meliefste K, Fischer P, Gehring U, et al. 2003. Estimating long-term average particulate air pollution concentrations: application of traffic indicators and geographic information systems. Epidemiology 4:228–239.

Cyrys J, Heinrich J, Hoek G, Meliefste K, Lewne M, Gehring U, et al. 2003. Comparison between different traffic-related particle indicators: elemental carbon (EC), PM2.5 mass, and absorbance. J Expo Anal Environ Epidemiol 13:134–143.

Gehring U, Heinrich J, Krämer U, Grote V, Hochadel M, Sugiri D, et al. 2006. Long-term exposure to ambient air pollution and cardiopulmonary mortality in women. Epidemiology 17:545-551.

Hochadel M, Heinrich J, Gehring U, Morgenstern V, Kuhlbusch T, Link E, et al. 2006. Predicting long-term average concentrations of traffic-related air pollutants using GIS-based information. Atmospheric Environment 40:542–553

Hoek G, Meliefste K, Cyrys J, Lewne M, Bellander T, Brauer M, et al. 2002. Spatial variability of fine particle concentrations in three European areas. Atmospheric Environment 36:4077–4088.

Lahmann E. 1997. Belastung der Umweltmedien: Staub und Staubinhaltsstoffe. In: Handbuch der Umweltmedizin (Wichmann HE, Schlipkoter HW, Fülgaff G, eds). Landsberg/Lech: Ecomed-Verlag.

Marple V, Rubow KL, Turner W, Spengler JD. 1987. Low flow rate sharp cut impactors for indoor sampling: design and calibration. JAPCA 1987;37:1303–1307

Palmes ED, Gunnison AF, Dimattio J, Tomczyk C. 1976. Personal sampler for nitrogen dioxide. Am Ind Hyg Assoc J 37: 570–577.

Umweltbundesamt. 2004. Handbuch für Emissionsfaktoren des Straßenverkehrs, Version 2.1/ Feb. 2004 UBA Berlin (www.hbefa.net).

VDI 3782 Bl. 7. 2002. Umweltmeteorologie; Kfz-Emissionsbestimmung.

Supplemental Material, Table 1. Comparison of study population and non-responder populations (SALIA study).

Variable	Participation in 2006 (n = 1775)		Death before 2006 (n = 585)		No participation in 2006 (n = 1933)		p
	Mean/proportion	N	Mean/proportion	N	Mean/proportion	N	
Age at baseline (years) [a]	54.4 (0.7)	1775	54.6 (0.5)	585	54.6 (0.5)	1932	<0.0001
C3c (mg/dl) [b]	91.5 (1.32)	1251	105.2 (1.35)	322	96.5 (1.33)	1122	0.0034
BMI (kg/m^2) [a c]	27.1 (4.1)	1602	28.7 (5.5)	406	28.0 (4.9)	1449	<0.0001
Hypertension	22.6%	1750	35.4%	573	28.0%	1889	<0.0001
Education <10yrs [d]	20.6%	1775	38.3%	585	36.6%	1933	<0.0001

[a] Mean (standard deviation).
[b] Geometric mean (geometric standard deviation).
[c] Measurements at day of baseline investigation, all others are questionnaire based results.
[d] Maximum number of years of education of study participant or husband.

Supplemental Material, Table 2. Association between incidence of type 2 diabetes and the potential confounders in Cox regression analyses for estimating the association between incidence of type 2 diabetes and exposure to ambient air pollution (as shown in Table 3 of the main manuscript)

Variable	Mutually adjusted HR (95% CI) [a]	p
Age (per year)	1.05 (1.00-1.10)	0.07
BMI (per 5 kg/m²)	1.98 (1.74-2.25)	<0.01
Education (<10 years vs ≥10 years)	1.08 (0.77-1.52)	0.65
Single room heating with fossil fuels	1.60 (1.15-2.22)	<0.01
Workplace exposures		
- Dust/fumes	1.11 (0.69-1.78)	0.67
- Extreme temperatures	1.58 (0.93-2.68)	0.09
Smoking (compared to never-smokers without ETS exposure)		
- Never-smokers/ETS exposure	0.92 (0.65-1.29)	0.62
- Ex-smokers	1.59 (1.00-2. 51)	0.05
- Current smokers (per 15 pack-years)	1.05 (0.86-1.29)	0.64

[a] Also adjusted for PM_{10} (1985-1989) as marker of exposure; HRs for covariates vary only marginally when replacing PM_{10} (1985-1989) with any other of the PM or traffic-related exposure variables.